Protein Science (2000), *9*:1922–1929. Cambridge University Press. Printed in the USA.
Copyright © 2000 The Protein Society

Lipoylating and biotinylating enzymes contain a homologous catalytic module

PEDRO A. RECHE

DNAX Research Institute, Department of Molecular Biology, 901 California Avenue, Palo Alto, California 94304-1104

(Received May 17, 2000; Final Revision July 10, 2000; Accepted July 24, 2000)

Abstract

Biotin and lipoic acid moieties are the covalently attached coenzyme cofactors of several multicomponent enzyme complexes that catalyze key metabolic reactions. Attachment of these moieties to the biotinyl- and lipoyl-dependent enzymes is post-translationally catalyzed by specific biotinylating and lipoylating protein enzymes. In *Escherichia coli*, two different enzymes, LplA and LipB, catalyze independent pathways for the lipoylation of the relevant enzymes, whereas only one enzyme, the BirA protein, is responsible for all the biotinylation. Counterparts of the *E. coli* BirA, LplA, and LipB enzymes have been previously identified in many organisms, but homology among the three families has never been reported. Computational analysis based on PSI-BLAST profiles and secondary structure predictions indicates, however, that lipoylating and biotinylating enzymes are evolutionarily related protein families containing a homologous catalytic module. Sequence conservation among the three families is very poor, but a single lysine residue is strictly conserved in all of them, which, according to the available X-ray crystal structure of the *E. coli* BirA protein, is expected to contribute to the binding of lipoic acid in the LplA and LipB enzymes.

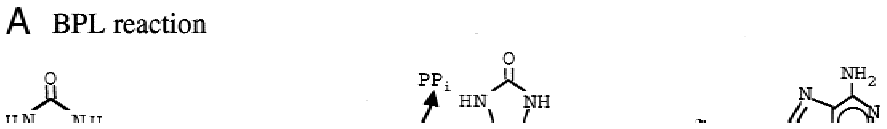
Keywords: biotinyl/lipoyl protein ligase; biotinylation/lipoylation; database searches; protein evolution; secondary structure; sequence space; structure prediction

Biotin and lipoic acid are the covalently bound cofactors of various multicomponent enzyme complexes that catalyze key metabolic reactions (Knowles, 1989; Perham, 1991). In these enzyme complexes, biotin and lipoic acid are attached via amide linkage through their carboxyl group and the ϵ-amino group of a specific lysine residue of a protein module known respectively as the biotinyl and the lipoyl domain. Biotinyl and lipoyl domains are structurally homologous proteins, with the respective biotinyl- and lipoyl-lysine residue prominently displayed at a tight β-turn of the structure. Thereby the biotinyl and lipoyl moieties shuttle catalytic intermediates between the successive active sites of the relevant enzyme complex (Perham & Reche, 1998). Covalent attachment of biotin and lipoic acid to these enzyme complexes occurs post-translationally, and it is mediated by biotinylating and lipoylating protein enzymes, which specifically recognize the biotinyl and lipoyl domains, ensuring their correct post-translational modification (Chapman-Smith & Cronan, 1999; Reche & Perham, 1999).

Reprint requests to: Pedro A. Reche, DNAX Research Institute, 901 California Avenue, Palo Alto, California 94304-1104; e-mail: reche@dnax.org.

Abbreviations: aaRS, aminoacyl-tRNA synthetase; ACP, acyl carrier protein; BPL, biotinyl protein ligase; LPL, lipoyl protein ligase (both LplA and LipB); LplA, lipoyl protein ligase A; LipB, lipoyl protein ligase B; LPT, lipoyl protein transferase; SS, secondary structure.

Biotin is attached to its target proteins by the action of biotinyl protein ligase (BPL), also known as biotin holoenzyme synthetase (Wood et al., 1980). This enzyme catalyzes the activation of biotin to biotinyl–AMP at the expense of ATP, and then transfers the biotinyl group to a specific lysine residue in the biotinyl domain (Fig. 1A). In *Escherichia coli*, biotinylation is catalyzed by the BirA protein, which also represses the transcription of the biotin operon, using the ligase reaction intermediate biotinyl-AMP as a co-repressor (Barker & Campbell, 1981; Cronan, 1989). The structure of *E. coli* BirA has been solved by X-ray crystallography (Wilson et al., 1992) and consists of three well-defined domains: DI, DII, and DIII (Fig. 2). DIII is at the C-terminus, comprises two three-stranded antiparallel β-sheets ordered in a SH3-like barrel fold and is of unknown function. DI occupies the N-terminus of BirA and encloses a DNA-binding helix-turn-helix motif. BirA binds to the biotin operon, repressing its transcription, via this domain (Buoncristiani et al., 1986; Wilson et al., 1992). Finally, the central domain DII, consisting of five α-helices (α1–α5) and a mixed β-sheet made up of seven strands (β1–β7), represents the BPL module of *E. coli* BirA and houses both the biotin and the ATP binding sites. The biotinyl moiety occupies a portion of the β-sheet face that is exposed to the solvent and makes contact with parts of three β-strands (β3, β6, and β7), the N-terminus of helix α2, and main-chain atoms of several residues at the loop connecting the strands β2 and β3 (see Fig. 2 for more comments). There is no direct crystallographic data for the ATP binding site,



but several lines of evidence indicate that the ATP binds adjacent to biotin and also that residues at the loop connecting the β2- and β3-strands contribute to the binding of ATP (Wilson et al., 1992). A region of homology to the BPL module of *E. coli* BirA can be identified in all biotinylating enzymes, while DI and DIII are only retained in some bacterial counterparts of the *E. coli* enzyme (this work).

Lipoylation, though similar to biotinylation, is more complicated. In *E. coli*, two independent lipoyl protein ligases (LPLs) encoded by the *lplA* and *lipB* genes have been reported (Morris et al., 1995). The lipoylation reaction catalyzed by LplA (Fig. 1B) is analogous to that of the biotinylation reaction mediated by the BPL, with lipoyl-AMP as the intermediate donor of the lipoyl moiety (Morris et al., 1994, 1995). In contrast, LipB (Fig. 1C) uses the lipoyl-acyl carrier protein (lipoyl-ACP) as the donor of the lipoyl group (Jordan & Cronan, 1997). ACPs are small proteins (40–70 residues) covalently modified in a serine residue with 4′-phosphopantetheine, and the lipoyl moiety is bound by thioester linkage to the 4′-phosphopantetheine cofactor (Jordan & Cronan, 1997). In humans and oxen, the encoding genes for lipoyltransferase (LPT) enzymes have been reported (Fujiwara et al., 1997, 1999). LPTs share extensive amino acid sequence similarity with *E. coli* LplA (Fujiwara et al., 1997, 1999), catalyzing also the transfer of lipoyl groups from lipoyl-AMP to the relevant apoproteins. However, unlike *E. coli* LplA, LPTs seem unable to synthesize the lipoyl-AMP intermediate from lipoyl and MgATP (Fujiwara et al., 1994).

LipB and LplA counterparts of the *E. coli* enzymes have been mentioned in various sources, but very little is known about them. Thus, no three-dimensional (3D) structures are available for any of the LPLs, and no structure–function studies have been carried out. In addition, homology between LplAs and LipBs has never been reported, and neither has homology between either of them and BPLs. In this paper, we show that the LplA and LipB enzymes

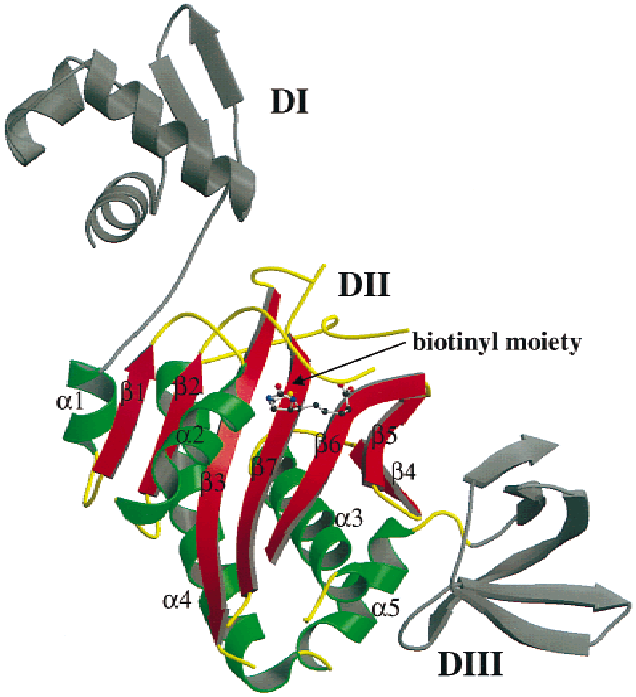


Fig. 2. 3D structure of the *E. coli* BPL. Ribbon drawing of the X-ray crystal structure of the *E. coli* BirA protein (Wilson et al., 1992). Domains DI, DII, and DIII are identified, and SS elements of the BPL module (DII) are labeled. The loops connecting the SS elements $\beta2$–$\beta3$, $\beta3$–$\alpha3$, $\beta6$–$\beta7$, and $\beta7$–$\alpha4$ do not appear in the structure because they were not visible in the electron density maps. Wilson et al. (1992) mentioned that the biotin binding site was located using byocitin (biotinyl–lysine) but it is free biotin that appears in the coordinates they deposited in the database (Protein Data Bank (PDB) entry 1bia). Judging from the deposited structure, the lysine moiety of biocytin was not visible on the electron density maps—probably because it was highly mobile—and therefore the authors preferred to model a carboxyl group ($-COO^-$) as in free biotin, instead of the amide linkage of biocytine (biotinyl-CO-NH-lysine).

contain a region with homology to the BPL module of biotinylating enzymes. This indicates that BPLs and LPLs (both LplAs and LipBs) are evolutionarily related protein families, with a homologous catalytic module that must have evolved from a common ancestor. Sequence conservation among the three families of protein ligases is poor, but a single lysine residue is retained in all sequences. Insights into evolutionary relationships among the three families will be discussed along with other implications derived from this unexpected relationship.

Results

Detection of homology between LPLs and BPLs

Homology between LplAs and LipBs (LPLs) has never been reported, and neither has homology between LPLs and BPLs, because their amino acid sequence similarities seem not to be significant. Surprisingly, in a search of the GeneBank nonredundant (NR) database using the program PSI-BLAST with the amino acid sequence of LPLA_MYCPN (GeneBank GI:2498521), several LipB enzymes came up after the first iteration with statistically significant scores (see Materials and methods for the PSI-BLAST settings). It was even more surprising to find that after the second iteration, the BPL of *Sinorhizobium meliloti* (GeneBank GI:6690793) was detected with a statistically significant e-value of 0.005. Searches initiated with LipB sequences also detected homology to BPLs, but more iterations were required. For example, a search initiated with LIPB_RICPR (GeneBank GI:6225619) detected the BPL of *Aquifex aeolicus* (GeneBank GI:2983199) with an e-value of 0.005 after the fifth iteration. Additional iterative searches using the PSI-BLAST program were carried out with various LplA and LipB sequences as query to propagate the detected homology to all members of the LPL and BPL families. As a rule, LplAs succeeded in detecting homology to BPLs only if they first succeeded in bringing LipB sequences into the equation. The same was true of LipB sequences: they detected homology to BPLs only if they first detected homology to LplAs. Also, homology to BPLs was detected much more frequently in searches started with LplA sequences than with LipB sequences, which in general converged without even detecting homology to LplAs. Searches initiated with BPL sequences converged before detecting homology to any LipB or LplA sequence, although frequently members of the two types of LPLs had the closest e-values to (albeit higher than) that of 0.01 fixed as a threshold. Thus, detection of homology between LPLs and BPLs is sequence dependent, and results from the combination of PSI-BLAST iterative profile searches with the use of statistically similar intermediate sequences that increase the detection of homology to remote homologues that are not directly related by statistically significant similarity (Holm, 1998; Park et al., 1997). It is also interesting to remark that detection of homology was always restricted to members of the LplA, LipB, and BPL families, and never reached any member of other protein families.

Detected sequence similarity between BPLs and LPLs spanned from the helix $\alpha4$ of the BPL module up to the region connecting the $\beta2$- and $\beta3$-strands, but secondary structure (SS) predictions helped to extend the similarity further to the N-terminus, up to the $\beta1$-strand of BPLs. A multiple sequence alignment with representative LplA, LipB, and BPL sequences is shown in Figure 3, and the derived evolutionary tree is shown in Figure 4. Also displayed in Figure 3 are the SS predictions of the aligned LPL and BPL sequences, which support the detected homology at a structural level.

Discussion

Structure comparison between the LPL and BPL catalytic modules

The $\alpha3$, $\beta1$, $\beta2$, $\beta3$, $\beta5$, $\beta6$, and $\beta7$ motifs of the BPL module are especially well defined in both LpLAs and LipBs (Fig. 3). Helix $\alpha1$ of the BPL module is substituted in LPLs by a predicted β-strand, and helix $\alpha2$ is predicted in LplAs but not in the LipB family (Fig. 3). An extra β-strand (βi-strand in Fig. 3) is predicted in the region connecting the $\beta2$- and $\beta3$-strands of LplAs and LipBs. This βi-strand is also predicted in BPLs, but it does not appear in the 3D structure of *E. coli* BPL (Fig. 2) because it is predicted in a region (residues 116–124) that was not visible on the electron density maps (Wilson et al., 1992). Interestingly, these observations, taken together, suggest that the region connecting the $\beta2$- and $\beta3$-strands is not totally unstructured, but highly mobile. The $\beta4$-strand of BPLs is predicted in the LplA family but not in LipBs. In the structure of *E. coli* BPL, the $\beta4$-strand encompasses only three residues, and though it is predicted in the aligned BPLs, its

signal is very weak (Fig. 3). Therefore, a β4-strand may yet appear in an eventual 3D structure of a LipB enzyme. Finally, the α5-helix of the BPL module finds no equivalent in either LipBs or LplAs. Some LipB sequences end right after the predicted α4-helix, and no sequence similarity is found after that helix (this work, not shown). Amino acid sequence conservation between LplAs extends around 85 residues toward the C-terminus after the predicted helix α4 (not shown), but with no sequence similarity to either BPLs or LipBs. Moreover, SS predictions do show a β-strand after the α4-helix of LplAs (not shown). Additional helical content is strongly predicted in the insertion region connecting the β7-strand and the α4-helix of LplAs (not shown).

Using the *E. coli* BPL coordinates as template and the multiple alignment of Figure 3, the models of the *E. coli* LplA and LipB enzymes were generated by homology modeling (Fig. 5) (see Materials and methods for details). Modeled 3D structures of both *E. coli* LPLs are restricted to their region of homology with *E. coli* BPL (from strand β1 to helix α4), but SS predictions suggest (Fig. 3) that the eventual 3D structure of the catalytic module of these enzymes should also include an additional β-strand in place of the α1-helix of the BPL module. Also, a predicted helix that precedes the former β-strand, marking the beginning of the homology between LplAs and LipBs (not shown), might pack with their modeled catalytic modules. The LplA enzyme contains additional SS features beyond the α4-helix (not shown), which may also pack with its modeled catalytic module.

Sequence comparison between the LPL and BPL catalytic modules

Amino acid sequence conservation between the catalytic modules of BPLs and LPLs is very low, and mainly affects residues that in the *E. coli* BPL enzyme are important for the scaffold of the structure, such as those contributing to the hydrophobic core (shown in gray in Fig. 3). Minimal amino acid sequence conservation is found in regions connecting the SS elements of the catalytic module, where long insertions can take place (Fig. 3). Despite the poor overall sequence similarity, a single lysine residue residue is strictly conserved in all BPL, LplA, and LipB sequences (Fig. 3). In *E. coli* BPL, this lysine residue corresponds to Lys183, and its amino group binds to the carboxyl group at the end of the hydrogen–carbon tail of biotin (Wilson et al., 1992). However, because the biotin binding site of *E. coli* BPL was located using biotinyl-lysine instead of free biotin (see Fig. 2 for comments), it follows that Lys183 would specifically bind to the carbonyl oxygen of the carboxyl group of biotin. Given the strict conservation of this lysine residue, we shall expect its function to be retained as well. Therefore, in LplA and LipB enzymes, the equivalent lysine residue should in the same manner bind to the carbonyl oxygen at the end of the hydrogen–carbon tail of the lipoyl moiety. Full validation of this prediction will await the structure of LpLA and LipB, but clearly, the homology between BPLs and LPLs and the strict conservation of the lysine residue indicate that binding of the biotin and lipoyl moieties to their respective ligases is somewhat similar. In this context, it is interesting to point out that lipoic acid and biotin share some stereochemical properties (Green, 1975). Indeed, lipoic acid binds to avidin and antibiotin antibodies (Harmon, 1980; Yamamoto & Sekine, 1987), and conversely, biotin can, in principle, bind to antilipoic acid antibodies.

More structure–function relationships can be inferred from the homology of LPLs with *E. coli* BPL. In *E. coli* BPL, it is known that several residues at the loop connecting the strands β2 and β3 contribute to substrate binding (of both biotin and ATP) and/or catalysis (Wilson et al., 1992), which is consistent with their strong conservation among BPLs (Fig. 3). Patterns of conserved residues also appear in the equivalent loop of LplAs and LipBs (Fig. 3), which most likely are also relevant for catalysis and substrate binding in those enzymes. Moreover, differences between the three families in the patterns of residue conservation at the loop connecting the strands β2 and β3 should be associated with their substrate specificity and/or their differences in catalysis. The loop connecting the strands β2 and β3 of *E. coli* BPL houses the sequence GRGRRG (Fig. 3), which has been postulated to be a consensus sequence for ATP binding (Wilson et al., 1992). In LplAs, no GRGRRG sequence motif is found, although according to the *E. coli* paradigm, they are expected to bind ATP (Morris et al., 1994, 1995). Thus, LplAs must have an alternative set of residues contributing to the binding of ATP, and some of them should be among the conserved residues at the loop connecting the β2- and β3-strands.

The catalytic module of LPLs and BPLs: Evolutionary relationships

LPLs and BPLs should have evolved from a common ancestor because sequence similarity and SS predictions indicate that they contain a homologous catalytic module. According to the alignment of Figure 3 and the derived evolutionary tree (Fig. 4), LplA and LipB sequences are closer to each other than to BPLs. In fact, although sequence similarity among individual LplA and LipB sequences can be statistically insignificant; overall, there is a significant overlap between the sequence space of LplAs and LipBs (sequence space is by definition the collection of all possible sequences, and the sequence space of a particular protein family is a subset containing sequences related by statistically significant sequence similarity, around 25% identity). Overlap between the sequence spaces of LplAs and LipBs may have indeed promoted some annotation mistakes, such as that of the SwissProt entry LIPB_AERPE, which is not a LipB enzyme but a LplA (see alignment in Fig. 3). On the other hand, the BPL family seems to be evolutionarily equidistant from both families of LPLs, LplA, and LipB, as indicated also by the evolutionary tree in Figure 4.

In mammals, synthesis of the activated intermediate lipoyl-AMP has been reported to be catalyzed by a lipoate-activating enzyme (Tsunoda & Yasunobu, 1967), whereas the transference of the lipoic group from the activated intermediate to the relevant protein substrates is catalyzed by LPT (Fujiwara et al., 1997, 1999). LPTs differ from bacterial LplA in their inability to synthesize lipoyl-AMP from lipoic acid and MgATP (Fujiwara et al., 1994), but with regard to sequence similarity, the distinction between them is not that clear (Fig. 3 and Fig. 4). In fact, the extent of the homology between LPTs and LplAs would suggest that the reported inability of mammalian LPTs to synthesize lipoyl-AMP from ATP and lipoic acid is due to technical problems and would also explain why no sequence has ever been deposited for any mammalian lipoate-activating enzyme.

Distant evolutionary relationships

The core fold of the *E. coli* BPL comprising the seven β-strands is also found in the catalytic domain of asparagine synthetase (Nakatsu et al., 1998) and in the catalytic domain of Class II of aminoacyl-tRNA synthetases (aaRS) (Cusack et al., 1990; Arty-

1BIA
BPL
LplA
LipB
BIRA_ECOLI/66-256 GI:115015
BIRA_BACSU/71-264 GI:773349
BIRA_METHT/6-198 GI:2623050
BIRA_PYRAB/1-178 GI:5457580
BPL_PYRAB/62-227 GI:5457589
BIRA_PARDE/1-179 GI:231637
BIRA_HAEIN/66-238 GI:1168673
BIRA_TREPA/4-188 GI:3322635
BPL_YEAST/369-608 GI:6320060
BPL_SCHPO/341-555 GI:2276356
BPL_ARATH/99-291 GI:1688072
BPL_HUMAN/454-654 GI:1813424
LPLA_ECOLI/31-216 GI:432634
LPLA_BACSU/27-214 GI:2226243
YQHM_BACSU/33-166 GI:2634887
LIPB_AERPE/27-213 GI:6685607
LPLA_MYCPN/30-217 GI:2498521
LPLA_CHLPN/33-213 GI:3328936
LPLA_MYCGA/29-216 GI:1480705
LPLA_AQUAE/91-283 GI:2983965
LPLA_SCHPO/49-237 GI:2257523
LPLA_CELEG/37-225 GI:3874964
LPT_HUMAN/58-243 GI:4586380
LPT_BOX /58-243 GI:2780412
LPLB_ECOLI/12-185 GI:2507595
LIPB_RICPR/30-209 GI:6225619
LIPB_MYCLE/39-220 GI:3122358
LIPB_MYCTU/40-212 GI:6016509
LIPB_SYNY3/34-212 GI:3122377
LIPB_KLULA/115-318 GI:3122354
LIPB_HAEIN/33-209 GI:1170791
LIPB_CYACA/30-207 GI:3122356
LIPB_YEAST/110-312 GI:6323268
LIPB_SCHPO/44-219 GI:3122361
LIPB_ARATH/134-316 GI:3122357
JPRED BPL
JPRED LPLA
JPRED LIPB
JPRED ALL
α1
β1
α2
β2
β3
α3
β4
β5
β6
β7
α4

miuk et al., 1994) (Fig. 6). In addition, Class II of aaRSs and asparagine synthetase have a two-step reaction mechanism analogous to that of BPL. Thus, both aaRSs and asparagine first catalyze the ATP-dependent formation of an aminoacyl-AMP intermediate (aspartyl-AMP in the case of asparagine synthetase), and then transfer the activated aminoacyl moiety to an acceptor tRNA or ammonia, respectively (Cusack, 1993; Richards & Schuster, 1998). Structural and mechanistic resemblance of Class II of aaRSs, asparagine synthetases, and BPLs has already led to the suggestion of an evolutionary relationship between them (Artymiuk et al., 1994), and to the speculation that the common ancestor was an ATP-binding domain able to activate the carboxyl group of small metabolites (S. Cusak in reply to Artymiuk et al., 1994). LplAs catalyze the synthesis of the activated intermediate lipoyl-AMP from ATP and lipoic acid, and therefore, fit into the above evolutionary scenario. The LipB enzyme, however, appears to be an exception. Jordan and Cronan (1997) have shown that lipoylation catalyzed by *E. coli* LipB proceeds using lipoyl-acyl carrier protein (lipoyl-ACP) as the donor of lipoic acid in a reaction that does not require ATP. Jordan and Cronan (1997) have also argued that lipoyl acid synthesis may occur via ACP-bound intermediates, the first of which (octanoyl-ACP) is produced by the fatty acid synthetic pathway. ACPs are small proteins (40–70 residues) that carry acyl groups via thioester linkage to the 4′-phosphopantetheine sulfhydryl group and transfer those acyl groups in several synthetic pathways (Jordan & Cronan, 1997). Therefore, it seems that the LipB enzyme may have obviated the need for ATP by using lipoyl-ACP as an alternative activated form of lipoic acid. Given this evidence, prior speculations about the original function of the common ancestor are cast into doubt and, although they may still be correct if we consider that the LipB family simply lost the postulated ability to use ATP in the course of evolution, the mechanistic feature common to all these proteins is indeed the transferring of a metabolite with an activated carbonyl group to a second metabolite.

Despite their putative evolutionary relationship, sequence spaces of Class II of aaRSs and asparagine synthetase do not overlap with those of BPLs and LPLs (this work). This, together with the fact that folds of proteins tend to recur in nature (Holm & Sander, 1996), suggests that more protein sequences and families containing the seven β-strand core of BPL remain to be discovered. Therefore, an ensemble of BPL, LPL, Class II of aaRS, and asparagine synthetase sequences allied with Hidden Markof Models (Hughey & Krogh, 1996) searches should be of great value in detecting new related protein sequences that may eventually link the sequence space gap between these protein families (Holm, 1998).

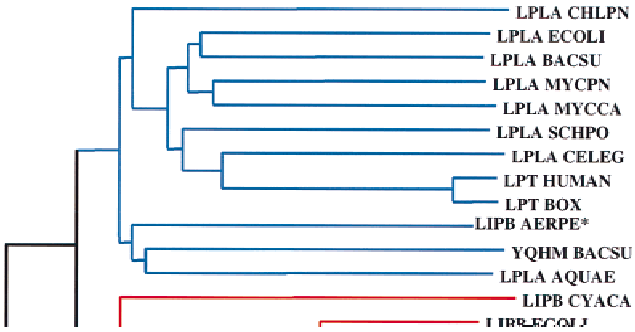


Fig. 3 (*facing page*).** Multiple sequence alignment of representative sequences of the BPL, LplA, and LipB families. Sequences are named following the SwissProt nomenclature (protein name_organism), and are identified by Genebank GI accession numbers. Only the region of homology between the three families of enzymes is shown. Extent of the aligned region is indicated for each sequence. Helix α1 of BPLs is also shown in the alignment; although in LPLs, it is replaced by a predicted β-strand. Insertion regions with no amino acid sequence similarity do not appear in the alignment, and the number of residues not shown is indicated in square brackets for each sequence. Sequences are arranged by families, and SS predictions for each family are shown under the aligned sequences (JPRED BPL, JPRED LPLA, and JPRED LIPB). A consensus prediction for all aligned sequences is also given (JPRED ALL). "E" is for β-strands and "H" for α-helices. SS motifs of the crystal structure of the *E. coli* BPL (PDB entry 1bia) are shown above its amino acid sequence. Cylinders are used to represent α-helices and arrows for β-strands. Residues that in the crystal structure of *E. coli* BPL contributed to the binding of biotinyl moiety of biocytin are indicated by an inverted triangle (▼). Shown framed is the loop region connecting the strands β2 and β3. This loop is partially unstructured or highly mobile in *E. coli* BPL (residues 116–124) and contributes to the binding of biotin and ATP (Wilson et al., 1992). Positions with identical residues or with a single residue change are shadowed in blue in the LplA family, in red in the LipB family, and in green in the BPL family. Positions occupied with residues that are identical in two families are shadowed in black. The Lys residue, strictly conserved in the three families, is shown in yellow font and shadowed in black. Positions shadowed in gray are occupied by hydrophobic residues in the three families, and in *E. coli* BPL contribute to the hydrophobic core of the structure (Wilson et al., 1992). The putative GRGRRG ATP binding sequence at the loop connecting the strands β2 and β3 of *E. coli* BPL is indicated with a red rectangle above the sequence. Note that the alignment reveals that the consensus ATP binding sequence in BPLs is GRGR rather than GRGRRG. Also note that the sequence in the SwissProt entry LIPB_AERPE belongs to the LplA family but not to the LipB family.

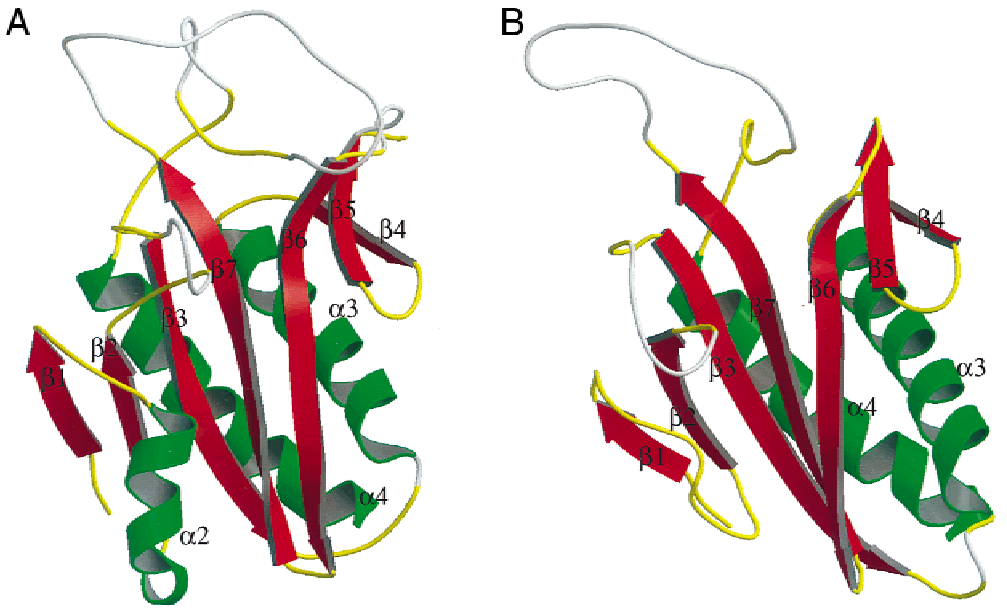


Fig. 5. Molecular models of LPLs. **A:** Ribbon drawing of the catalytic module of *E. coli* LplA. **B:** Ribbon drawing of the catalytic module of *E. coli* LipB. Both models were generated by homology modeling (see Materials and methods) using the coordinates of the BPL module from *E. coli* BirA as a template. Helix $\alpha 2$ was not modeled in the LipB enzyme because the SS algorithms did not predict it in LipB enzymes. Strand $\beta 4$ was not predicted in LipB enzymes but still was modeled for reasons explained in the text. Loops connecting the secondary structure elements $\beta 2$–$\beta 3$, $\beta 3$–$\alpha 3$, $\beta 6$–$\beta 7$, and $\beta 7$–$\alpha 4$ are not expected to be modeled properly because they are not visible in the *E. coli* BPL template (Wilson et al., 1992). Moreover, as discussed in the text, predictions indicate that additional SS features might appear in those loops.

Conclusions

BPLs and LPLs catalyze the post-translational attachment of the relevant biotin and lipoic acid moieties to a specific lysine residue located at the tip of the β-turn of their structurally homologous protein substrates, the biotinyl and lipoyl domains, respectively (Perham & Reche, 1998). The extraordinary ability of BPLs and LPLs to distinguish between their homologous substrates was thought to be facilitated if, as sequence similarity apparently indicated, BPLs and LPLs were structurally different enzymes (Reche & Perham, 1999). However, this work supports that LPLs and BPLs contain a homologous catalytic domain. This unexpected relationship is of special interest because no 3D structure is available for any LPL, not are there any structure–function studies. Thus, the homology between LPLs and *E. coli* BPL, whose structure is known (Wilson et al., 1992), has opened the possibility of identifying residues and regions that might be involved in substrate binding and/or catalysis in LPLs, most notably represented by a strictly conserved Lys residue that is expected to bind to the carboxy–carbonyl group of the lipoyl moiety.

Materials and methods

Database searches and sequence alignments

Sequence similarity searches were carried out using the PSI-BLAST program (Altschul et al., 1997) on the NR database of the National Center for Biotechnology Information (NCBI). Briefly, this program constructs a position-dependent weight matrix (profile) from multiple alignments of BLAST hits that are above the expectation value (e-value) and then iterates the search using this evolving profile as query. PSI-BLAST was run to convergence, with the e-value 0.01 used as the cutoff, and a Blosum 45 matrix. Multiple sequence alignments for the LplA and LipB subfamilies were constructed with the program ClustalX1.8 and edited manually to adjust PSI-BLAST local alignments. BPL multiple sequence alignment was initially obtained from the PFAM database (www.sanger.ac.uk/Software/Pfam/) and then corrected manually, guided by the structure of *E. coli* BPL to remove gaps occurring in secondary structure elements. LplA, LipB, and BPL subfamilies were aligned with each other with the program ClustalX1.8 used under the profile mode, and the resulting alignment was adjusted to match PSI-BLAST local alignments and secondary structure predictions.

Other procedures

Protein secondary structure predictions were carried out on the JPRED server (http://jura.ebi.ac.uk:8888/), which gives a consensus secondary structure prediction based on different algorithms. Predictions were obtained using multiple sequence alignments as input. Evolutionary trees were obtained from alignments using the Neighbor-joining (Firestine et al., 1996) method implemented in the ClustalX1.8 program. Molecular 3D models of the *E. coli* LipB and LplA enzymes were obtained by means of homology modeling, using the Modeller package (Sâli & Blundell, 1993) and the coordinates of the BPL module of *E. coli* BirA (Wilson et al., 1992). The alignment between template and the relevant sequences was obtained as indicated elsewhere. Models were optimized using the command OPTIMIZE of the Modeller package. Ribbon representations of the structures were generated using MOLSCRIPT (Kraulis, 1991), followed by image rendering with Raster3D (Merritt & Murphy, 1994). Schematic topologies of 3D protein structures were obtained from the Atlas of Protein Topology Cartoons (http://tops.ebi.ac.uk/tops/AtlasHTML.html) (Westhead et al., 1999).

Acknowledgments

The author wishes to acknowledge Dr. A. Sâli for permission to use the program modeller; Dr. J. Fernando Bazan for support and comments; Drs. Gerard Zurawski and Robert Kastelein for corrections on the manuscript; and Dr. Dovie R. Wylie for editorial assistance. Research at DNAX is supported by Schering-Plough.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res 25*:3389–3402.

Artymiuk PJ, Rice DW, Poirrette AR, Willet P. 1994. A tale of two synthetases. *Nat Struct Biol 11*:758–760.

Barker DF, Campbell AM. 1981. Genetic and biochemical characterization of the *birA* gene and its product: Evidence for a direct role of biotin holoenzyme synthetase in repression of the biotin operon in *Escherichia coli*. *J Mol Biol 146*:469–492.

Buoncristiani MR, Howard PK, Otsuka AJ. 1986. DNA-binding and enzymatic domains of the bifunctional biotin operon repressor (BirA) of *Escherichia coli*. *Gene 44*:255–661.

Chapman-Smith A, Cronan JE Jr. 1999. The enzymatic biotinylation of proteins: A post-translational modification of exceptional specificity. *Trends Biochem Sci 24*:359–363.

Cronan JE Jr. 1989. The *E. coli bio* operon: Transcriptional repression by an essential protein modification enzyme. *Cell 58*:427–428.

Cusack S. 1993. Sequence, structure and evolutionary relationships between class 2 aminoacyl-tRNA synthetases: An update. *Biochimie 75*:1077–1081.

Cusack S, Berthet-Colominas C, Hartlein M, Nassar N, Leberman R. 1990. A second class of synthetase structure revealed by X-ray analysis of *Escherichia coli* seryl-tRNA synthetase at 2.5 Å. *Nature 347*:249–255.

Firestine SM, Nixon AE, Benkovic SJ. 1996. Threading your way to protein function. *Chem Biol 3*:779–783.

Fujiwara K, Okamura-Ikeda K, Motokawa Y. 1994. Purification and characterization of lipoyl-AMP:N^{ϵ}-lysine lipoyltransferase from bovine liver mitochondria. *J Biol Chem 269*:16605–16609.

Fujiwara K, Okamura-Ikeda K, Motokawa Y. 1997. Cloning and expression of a cDNA encoding bovine Lipoyltransferase. *J Biol Chem 272*:31974–31978.

Fujiwara K, Suzuki M, Okumachi Y, Okamura-Ikeda K, Fujiwara T, Takahashi E, Motokawa Y. 1999. Molecular cloning, structural characterization and chromosomal localization of human lipoyltransferase gene. *Eur J Biochem 260*:761–767.

Green NM. 1975. Avidin. *Adv Protein Chem 29*:85–133.

Jordan SW, Cronan JE Jr. 1997. A new metabolic link. The acyl carrier protein of lipid synthesis donates lipoic acid to the pyruvate dehydrogenase complex in *Escherichia coli* and mitochondria. *J Biol Chem 272*:17903–17906.

Harmon FR. 1980. Purification of antibodies against biotin of lipoic acid-Sepharose. *Anal Biochem 103*:58–63.

Holm L. 1998. Unification of protein families. *Curr Opin Struct Biol 8*:372–379.

Holm L, Sander C. 1996. Mapping the protein universe. *Science 273*:595–603.

Hughey R, Krogh A. 1996. Hidden Markov models for sequence analysis: Extension and analysis of the basic method. *Comput Appl Biosci 12*:95–107.

Knowles JR. 1989. The mechanism of biotin-dependent enzymes. *Annu Rev Biochem 58*:195–221.

Kraulis PJ. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr 24*:946–950.

Merritt EA, Murphy MEP. 1994. Raster3D Version 2.0: A program for photorealistic molecular graphics. *Acta Crystallogr D50*:869–873.

Morris TW, Reed KE, Cronan JE Jr. 1994. Identification of the gene encoding lipoate-protein ligase A of *Escherichia coli*. Molecular cloning and characterization of the *lplA* gene and gene product. *J Biol Chem 269*:16091–16100.

Morris TW, Reed KE, Cronan JE Jr. 1995. Lipoic acid metabolism in *Escherichia coli*. The *LplA* and *LipB* genes define redundant pathways for ligation of lipoyl groups to apoprotein. *J Bacteriol 177*:1–10.

Nakatsu T, Kato H, Oda J. 1998. Crystal structure of asparagine synthetase reveals a close evolutionary relationship to class II aminoacyl-tRNA synthetase. *Nat Struct Biol 5*:15–19.

Park J, Teichmann SA, Hubbard T, Clothia C. 1997. Intermediate sequences increase the detection of homology between sequences. *J Mol Biol 273*: 349–353.

Perham RN. 1991. Domains, motifs and linkers in 2-oxo acid dehydrogenase multienzyme complexes: A paradigm in the design of a multifunctional protein. *Biochemistry 30*:8501–8512.

Perham RN, Reche PA. 1998. Swinging arms in multifunctional enzymes and the specificity of post-translational modification. *Biochem Soc Trans 26*: 299–303.

Reche P, Perham RN. 1999. Structure and selectivity in post-translational modification: Attaching the biotinyl-lysine and lipoyl-lysine swinging arms in multifunctional enzymes. *EMBO J 18*:2673–2682.

Richards NG, Schuster SM. 1998. Mechanistic issues in asparagine synthetase catalysis. *Adv Enzymol Relat Areas Mol Biol 72*:145–198.

Sâli A, Blundell TL. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol 234*:779–815.

Tsunoda JN, Yasunobu KT. 1967. Mammalian lipoic acid activating enzyme. *Arch Biochem Biophys 118*:395–401.

Westhead DR, Slidel TW, Flores TP, Thornton JM. 1999. Protein structural topology: Automated analysis and diagrammatic representation. *Protein Sci 4*:897–904.

Wilson KP, Shewchuk LM, Brennan RG, Otsuka AJ, Matthews BW. 1992. *Escherichia coli* biotin holoenzyme synthase/*bio* repressor crystal structure